



King's Research Portal

DOI:

[10.1109/TMI.2011.2162529](https://doi.org/10.1109/TMI.2011.2162529)

Document Version

Early version, also known as pre-print

[Link to publication record in King's Research Portal](#)

Citation for published version (APA):

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Rutherford, M., Edwards, A., Hajnal, J., & Rueckert, D. (2011). A Combined Manifold Learning Analysis of Shape and Appearance to Characterize Neonatal Brain Development. *Ieee Transactions on Medical Imaging*, 30(12), 2072-2086. [5958609].
<https://doi.org/10.1109/TMI.2011.2162529>

Citing this paper

Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights

Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the Research Portal

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

A combined manifold learning analysis of shape and appearance to characterize neonatal brain development

P. Aljabar^{1*}, R. Wolz¹, L. Srinivasan², S.J. Counsell³, M.A. Rutherford³,
A.D. Edwards², J.V. Hajnal³ and D. Rueckert¹

Abstract

Large medical image datasets form a rich source of anatomical descriptions for research into pathology and clinical biomarkers. Many features may be extracted from data such as MR images to provide, through manifold learning methods, new representations of the population's anatomy. However, the ability of any individual feature to fully capture all aspects morphology is limited. We propose a framework for deriving a representation from multiple features or measures which can be chosen to suit the application and are processed using separate manifold-learning steps. The results are then combined to give a single set of embedding coordinates for the data. We illustrate the framework in a population study of neonatal brain MR images and show how consistent representations, correlating well with clinical data, are given by measures of shape and of appearance. These particular measures were chosen as the developing neonatal brain undergoes rapid changes in shape and MR appearance and were derived from extracted cortical surfaces, non-rigid deformations and image similarities. Combined single embeddings show improved correlations demonstrating their benefit for further studies such as identifying patterns in the trajectories of brain development. The results also suggest a lasting effect of age at birth on brain morphology, coinciding with previous clinical studies.

Index Terms

¹Biomedical Image Analysis Group, Department of Computing, Imperial College London, London, SW7 2AZ, UK.

²Division of Neonatology, Imperial College Healthcare NHS Trust, London, UK

³Institute of Clinical Sciences, Imperial College London and MRC Clinical Sciences Centre, Hammersmith Hospital, London, UK

*Correspondence: paul.aljabar@imperial.ac.uk

MR Images, Manifold Learning, Dimensionality reduction, Neonatal brain development

I. INTRODUCTION

Large scale imaging studies have become an established approach for supporting biological and clinical research in general and for the discovery of biomarkers in particular. Neuroscientific and neuroclinical applications form a significant part of this research largely thanks to the increasingly wider availability of large databases of brain image data, especially those containing structural magnetic resonance (MR) images.

An established approach to such large scale studies is to identify a particular feature of interest in each image, for example the grey matter density at a specific location in the brain, and subsequently to measure this feature across the population. Studies of pathology or searches for biomarkers may then be carried out by, for example, separating the sample into patients and healthy controls and seeking associations between the disease and the chosen feature. Studies using voxel-based morphometry (VBM) or deformation-based morphometry (DBM) form good examples of this approach [1]–[6]

An alternative paradigm has begun to be applied to neuroimaging studies using methods recently developed in multivariate statistics, machine learning and pattern recognition [7]–[11]. This is based on a class of techniques developed to generate useful and low dimensional representations of high dimensional data. The techniques are described collectively as methods of manifold learning or dimensionality reduction and a good overview may be found in [12].

Medical images can be very high dimensional, around a million voxels in an MR image is typical, and therefore they are suitable data for such dimensionality reduction or manifold learning approaches. This is based on the intuition that each image of a region, such as a brain, may be viewed as a single point in a high-dimensional space but that the set of all such points is unlikely to be uniformly distributed throughout this space. Rather, they postulated to lie on or near an embedded sub-manifold with a much lower dimension and the new representation sought is a low-dimensional coordinate parametrization of such a manifold.

The image data may be used directly to learn a set of embedding coordinates but an emerging set of powerful techniques adopt the alternative approach of using measures of proximity which relate *pairs* of data items. In this way, the pairwise measures may be chosen to reflect specific aspects of the data that are of interest for the analysis and relevant to the population being studied.

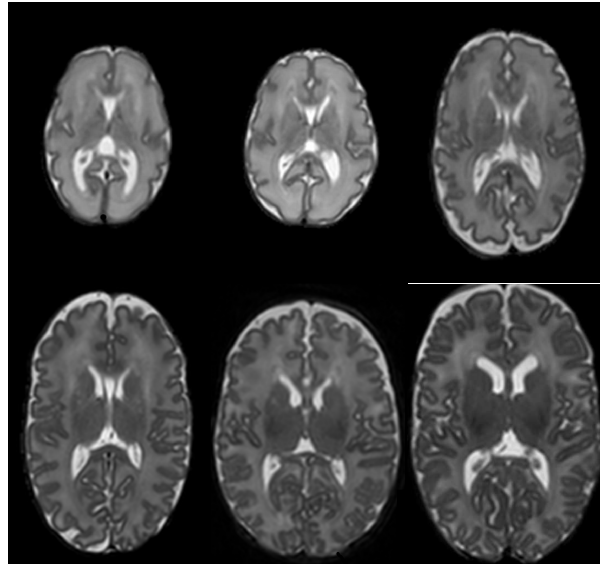


Fig. 1. Apart from the overall change in size, the neonatal brain undergoes significant macroscopic changes in shape (for example in cortical folding in the relative size of the ventricles) and in appearance (for example in relative contrasts across tissues) due to processes at the cellular scale. The subjects above were scanned at 29, 31, 33, 36, 40 and 44 weeks gestational age (See also Figure 2).

Langs et al. [13] use model description length (MDL) to assess pairwise similarities among point-sets which they use in an application of diffusion maps [14]. The derived representations can be used to group points in images into coherent clusters in the form of ‘shape maps’. These can be incorporated into medical imaging applications that rely on landmark based analysis. Coordinate embeddings derived from Laplacian eigenmaps [15] were used by Aljabar et al. [16] to help discriminate between Alzheimer’s Disease (AD) patients and healthy matched controls. In this study, pairwise measures of similarity were used which were derived from overlaps of labels in structural segmentations of the image data.

In contrast, a distance measure (rather than a similarity) was defined between pairs of images by Gerber et al. [17] and analysed using the Isomap manifold learning algorithm [18]. The distances were derived from non-rigid transformations and were used to derive embedding coordinates for a manifold structure representation of a set of brain MR images of AD patients and controls. Deformation derived distances were also used to approximate a manifold structure for brain images by Hamm et al. [19]. In this case the structure can be viewed as a k-nearest neighbour graph where nodes represent images and edges only join pairs of similar images (i.e.

separated by a small distance). By navigating along edges it is possible to estimate correspondences between highly dissimilar images as a composition of relatively small transformations between similar image pairs.

Wolz et al. [20] also used the navigation of local neighbourhoods in a low-dimensional embedding in order to construct atlas-based segmentations in a large and anatomically heterogeneous image dataset comprising patients with dementia and matched controls. The approach produced accurate segmentations even for images that were very dissimilar from the manually constructed atlases. A manifold learning approach was applied by Sun et al. [21] to derive a low dimensional representation for patterns of cortical folding. In this study, the distance between each pair of 62 subjects was obtained from pre-extracted outlines of their cortical sulci and the Isomap algorithm was used to generate the low dimensional embedding coordinates. A further clustering step was then used to identify characteristic sub-groups of the population. By focusing on different sulci, this work demonstrates a regional strategy to deriving the embedding.

More recent studies on medical data include work on manifold learning for endoscopic video data [22], for gating ultrasound data of the chest to account for breathing [23] and for classifying morphometry [24]. In the broader context of image processing, manifold learning methods have been used for recognition tasks [25], [26] or, for example, to regress the age of subjects from face images [27].

The type of analysis described above can be viewed as relying upon or generating features in the image data. The initial pairwise measures, whether of distance or similarity, can be viewed as first stage features and the derived embedding coordinates can be viewed as second stage features derived from the first. The choice of which pairwise measure to use depends on the particular application and should reflect those aspects of the morphology which are relevant to the application and the population being studied. It is unlikely, however, that starting with any individual pairwise measure among the images will lead to a complete characterization of the data.

We therefore contribute a framework where multiple pairwise measures are used in separate manifold learning steps and the resulting embedding coordinates are fused to produce a single combined representation. By using different pairwise measures we can focus on different aspects of the morphology during the analysis; those aspects that are relevant to the underlying biological processes or systems being studied.

To demonstrate and motivate the approach we present an example application of a population study of the developing brain using a set of neonatal brain MR images acquired at a range of ages. The pairwise measures used are selected to reflect the brain’s shape and appearance as they both undergo significant change during early neonatal development due to growth at the macroscopic scale, for example in cortical folding, and to the complementary changes in MR signal due to microscopic structural processes such as myelination (See Figures 1 and 2.)

Neuroimaging studies of the neonatal and fetal brain face different challenges from those typically encountered in studies of adult MR data. As well as an increase in variation due to comparatively rapid morphological change, imaging the perinatal brain is generally more difficult due to the increased chance of motion artefacts and the varying MR signal response of different tissues with age. This has led to a number of studies in which traditional neuroimaging tasks have been adapted to specifically address perinatal data. A common characterisation of the developing brain is obtained through spatio-temporal atlases of cross-sectional data acquired at different ages. Good examples are provided by Habas et al. [28] and Murgasova et al. [29] for fetal and neonatal data respectively. Adaptations of segmentation techniques to neonatal and fetal data have also been developed in the contexts of longitudinal analysis of neonatal and fetal data [30], [31] and of multi-region atlas-based segmentation [10]. Models for the patterns of MR signal change in neonatal tissue have also been developed [32], [33]. The work presented in this paper can be viewed as generating new representations for image data which may subsequently be used in the types of analysis described above. As well as the typical neonatal studies described above, the representations for image sets given by the proposed framework may also be applied in the context of other more general studies which seek to find ‘modes’ or exemplars of a population of images ([34] or [35] for example). It should be noted, however, that while the example application focuses on neonatal brain image data, the proposed framework for combining information from multiple measures may be applied more generally to other populations and other forms of imaging data.

In the exemplar application presented, a comparison against non-imaging clinical information relating to the subjects, such as the age of the subjects at scan or their ages at birth provides consistent and high correlations with the separate shape- and appearance-based embeddings. We show that fusing the embeddings leads to improved correlations with non-imaging clinical data and provides a useful new parametrization of the images that may be used, for example, to

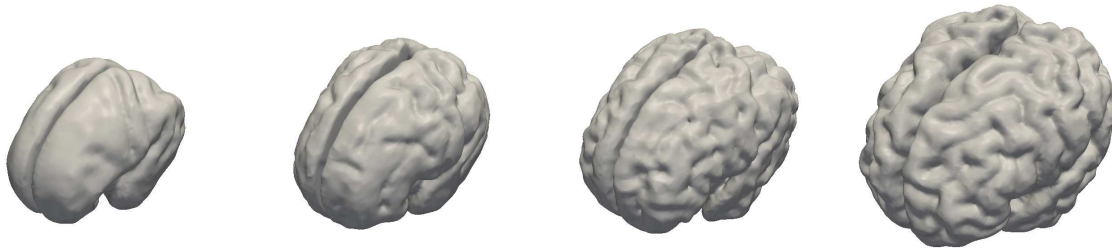


Fig. 2. An illustration of the change in the cortex with age for infants scanned at 29, 32, 36 and 39 weeks (See also Figure 1).

represent trajectories of change or for extracting biomarkers of development during this important stage of brain growth.

This submission builds on the work previously described in [36] and in this work we further seek to determine the best combination of methods by which the heterogeneous pairwise measures may be combined into a single unified manifold representation. The pairwise measures are values derived from either similarity or distance functions and different manifold learning models assume their input is one or the other of these types of measure. This can necessitate the conversion of the data type (e.g. from distances to similarities) prior to analysis. Using an exemplar manifold learning model for similarity data and one for distances, we use a rank correlation based analysis to assess all combinations of data and manifold learning technique to select the optimal method combination to be used within the framework. As well as inter-subject measures we previously derived from image intensities and deformations [36], we demonstrate the extensibility of the framework by incorporating further shape information through pairwise comparisons of cortical surface representations for the subjects. This is based on the use of the ‘currents’ metric originally developed for surface matching [37] and is the first application of this measure which we are aware of outside the context of registration. We also compare early- and late-combination strategies for the information derived from the different measures studied. This is either through the merger of measures before a single embedding step or through the combination of multiple coordinate sets after separate embedding steps.

II. METHODS

We focus on manifold learning methods which rely on pairwise measures among images and we use the image intensities, deformations between them, and comparisons of cortical surfaces to

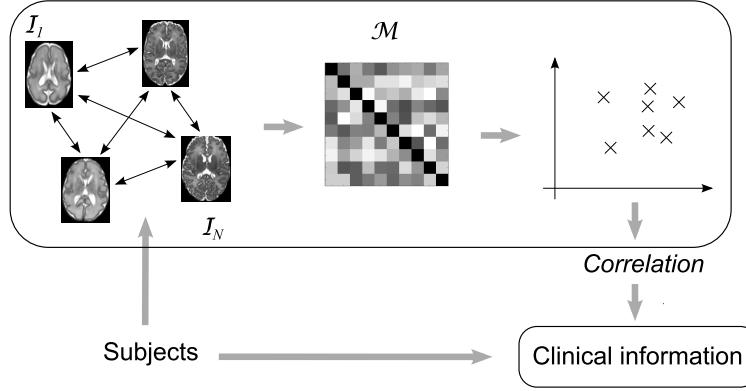


Fig. 3. Manifold learning for populations studies of image data. All subjects' images are registered in pairs (top left). A chosen measure \mathcal{M} is estimated for each pair to create a weights matrix (top middle) which is used to create a set of embedding coordinates with a chosen number of dimensions (top right where each point represents an image). Meaningful associations between embedding coordinates and clinical information indicate the representational power of the embedding.

provide these measures. A schematic illustration of the way in which a manifold learning method can be used within a populations study is given in Figure 3. In particular, we show how multiple pairwise measures may be used, each aimed at a feature of the data and providing its own set of embedding coordinates. In our example application, measures are chosen to reflect shape and appearance in neonatal brain images. The separate embeddings produced are then fused after scaling, concatenation and a further dimension reduction step. We show that this leads to a single combined embedding that relates more strongly to clinical information indicating that it may serve as a better representation of the image data. An overview diagram of the steps used in the framework is given in Figure 4.

A. Images, registrations and pairwise measures.

Let $\{\mathcal{I}_i : i = 1, \dots, N\}$ be a collection of images defined on a discrete domain $\Omega \in \mathbb{R}^3$. For simplicity, each \mathcal{I}_i is assumed to be a scalar valued function, i.e. $\mathcal{I}_i : \Omega \rightarrow \mathbb{R}$. For a given pair of images, \mathcal{I}_i and \mathcal{I}_j , and assuming \mathcal{I}_i is the fixed or target image, a registration algorithm may be used to define the correspondence between them via a transformation $\phi_{ij} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ which optimises some energy or image similarity function \mathcal{E} :

$$\phi_{ij} = \underset{\phi}{\operatorname{argmin}} \mathcal{E}(\mathcal{I}_i, \mathcal{I}_j \circ \phi).$$

Assuming that ϕ_{ij} are available for $i \neq j$ and that $\phi_{ii}(x) = x$ is the identity transformation, it becomes possible to derive measures of proximity between all pairs of images $\{\mathcal{I}_i, \mathcal{I}_j\}$ using either their intensity data directly or using features extracted from the image data. The measures of proximity may be general and can represent similarities, distances or discrepancies between each pair. In this work we apply the framework to carry out a population study based on images of the neonatal brain and we choose correspondingly suitable pairwise measures: An appearance measure derived from the intensity similarity of the images $I_i(x)$ and $I_j(\phi_{ij}(x))$ and shape measures derived either from a norm on the transformation ϕ_{ij} or from a comparison of cortical meshes extracted from the images. In general, the number and choices of pairwise measures need to be chosen to suit the application.

B. Measures used for the neonatal data

The measures used in the example application were defined as either shape or appearance measures. The appearance measure was derived from the image intensity data and two shape measures were used, derived either from deformations between images or from a distance based on a currents metric between cortical representations [37].

We define the appearance measure, \mathcal{M}^a , to be the normalised cross correlation of the images:

$$\mathcal{M}^a(\mathcal{I}_i, \mathcal{I}_j; \phi_{ij}) = \frac{1}{\sigma_i \sigma_j |\Omega|} \sum_{x \in \Omega} (\mathcal{I}_i(x) - \mu_i)(\mathcal{I}_j(\phi_{ij}(x)) - \mu_j) \quad (1)$$

where μ_i , μ_j , σ_i and σ_j are the means and standard deviations of I_i and $I_j \circ \phi_{ij}$. Other image intensity-based similarity measures are possible such as mutual information, etc.

The deformation shape measure, \mathcal{M}^{sd} , is defined to represent the degree of deformation required to transform one image into another. This draws on approaches which derive metrics in the setting of diffeomorphic transformations, such as the work of Beg et al. [38], in which $\phi(x)$ is typically defined as the end point at $t = 1$ of a streamline obtained by integrating a time-dependent velocity field $v(x, t) : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$. In particular $\phi(x, 0) = x$, $\phi(x, 1) = \phi(x)$ and

$$\frac{d}{dt}\phi(x, t) = v(\phi(x, t), t)$$

In practice, the transformation of a point may be estimated by a numeric implementation of Euler integration. As a concrete example, if the time interval $[0, 1]$ is divided into two steps,

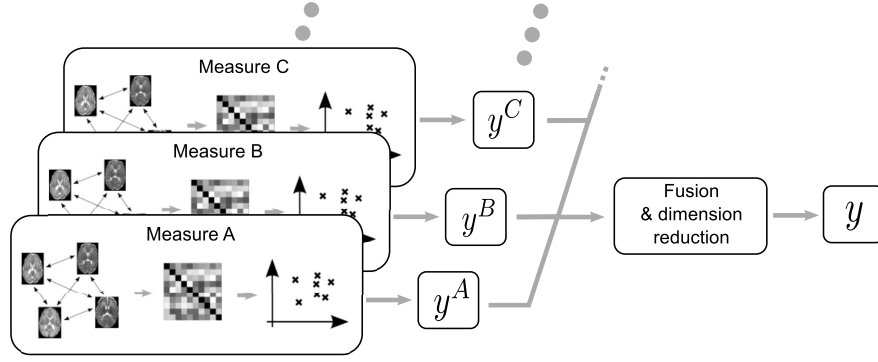


Fig. 4. An overview of the framework for combining multiple measures through manifold learning. Separate measures of interest are used to generate multiple embeddings. The results are fused into a single embedding that provides an improved representation of the original data. In our example, measures of shape and appearance were used although the type and number of measures may be varied. Additionally, the dimension reduction method(s) used in the intermediate embeddings and in the final combination step may be chosen independently.

then one may use the approximation

$$\phi(x, 0) = x$$

$$\phi(x, 0.5) \approx x + 0.5v(x, 0)$$

$$\phi(x) = \phi(x, 1) \approx x + 0.5v(x, 0) + 0.5v(\phi(x, 0.5), 0.5)$$

In this model, the transformation is effectively parametrized through its velocity field which is optimised so that the images are aligned. This problem is ill-posed with multiple solutions so the choice of v is made so that an energy term equal to

$$\int \int_0^1 \|v(x, t)\|_L^2 dt dx$$

is minimised for a given choice of differential operator L . The minimal value is defined as the value of metric induced by ϕ to describe the distance between the images.

In this work, we follow the example of Gerber et al. [17] and Hamm et al. [19] who simplify the estimation of the deformation metric by using a single displacement field $u : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}^3$ to represent ϕ in a small deformation setting to give $\phi(x) = x + u(x)$. This is equivalent to assuming a single discrete time step in the diffeomorphic formulation above. In this case, the

pairwise shape measure between images \mathcal{I}_i and \mathcal{I}_j may be similarly defined as \mathcal{M}^{sd} where

$$\mathcal{M}^{sd}(\mathcal{I}_i, \mathcal{I}_j; \phi_{ij}) = \sum_{x \in \Omega} \|u_{ij}(x)\|_L^2 dx \quad (2)$$

Local structure is important when learning a low-dimensional representation. For example, all pairwise measures are used to construct a weights matrix in the analysis and this is typically sparsified to represent a k-NN graph before use. The intuition in sparsifying the weights matrix, and the corresponding graph, can be viewed as modelling the reliability of pairwise measures. If the items are ‘close’ or ‘similar’, the pairwise measure is assumed to be more reliable and more useful for further processing than measures between highly distinct items (For further details, see Section II-D, Equation 6 or [15], [18], [39] for example). This aspect of learning embeddings makes the use of a small deformation setting for registration appropriate as the measures among more similar images (corresponding to distances at a local scale) are the ones which will characterize the final representation. The summed norm in Equation 2 is calculated as $\|u_{ij}(x)\|_L^2 = \|Lu_{ij}(x)\|_2^2$ and we set $L = I_3 + \alpha \nabla$, where I_3 is the 3×3 identity matrix and ∇ denotes the divergence operator. Using notation similar to that in [17], this means that the distance for a displacement field u is a combination of the total displacement of the deformation, $\|u(x)\|_2^2$ and its first order derivatives $\|\nabla u\|_2^2$. The weighting factor α controlling their relative contributions is empirically set to 0.1 for all experiments in the exemplar application presented.

The pairwise measure of shape based on cortical surfaces, \mathcal{M}^{sc} , was derived using the currents-based metric which has previously been defined and used in the context of surface registration [37] or in the incorporation of surface information within an image registration framework [40].

The cortical boundaries in the images were estimated using the method described in [41] which uses a tissue segmentation of the neonatal images to initialise a level set based estimation of the cortical boundary. This provides an implicit representation of the cortex as a distance map and the marching cubes algorithm was used to generate a representation of the zero isosurface as a triangulated mesh. Example renderings of cortical surfaces extracted at different ages are given in Figure 2. For an image \mathcal{I} , the corresponding cortical mesh, \mathcal{C} consists of a set of K faces $\{f_k : k = 1 \dots K\}$. The current representation can be approximated using the centroid location of each face c_k and its normal vector η_k with magnitude equal to the face area. The collection $\{(c_k, \eta_k) : k = 1 \dots K\}$ gives the currents-based representation of the surface as a set

of vector valued Dirac masses (see [40], Eqn. 3).

The norm of a surface in this formulation is given by a weighted sum of the inner products of pairs of normals

$$\|\mathcal{C}\| = \sum_{k=1}^K \sum_{k'=1}^K \eta_k \cdot \eta_{k'} \kappa(c_k, c_{k'}) \quad (3)$$

where the weights are given by κ , a symmetric kernel function that decreases with the distance between the centres. In this work we simply use a Gaussian kernel with a given width λ which controls its scale, i.e.

$$\kappa(c, c') = \exp(-(c - c')^2 / \lambda^2) / \lambda \sqrt{2\pi}.$$

The currents formulation as a collection of centre/normal pairs allows the negation of a surface, ‘ $-\mathcal{C}$ ’, to be represented by negating the sense of its normal vectors, i.e. $\eta_k \leftarrow -\eta_k, \forall k$. The union of a pair of surfaces is easily represented by the union of their collections of centre/normal pairs. The current based distance between a pair of surfaces \mathcal{C}_1 and \mathcal{C}_2 , $\|\mathcal{C}_1 - \mathcal{C}_2\|_c$ is defined as the norm of $\mathcal{C}_1 \cup \{-\mathcal{C}_2\}$ using Equation 3. This means that, with a suitable indexing on the centres and normals, the distance between the surfaces may be written

$$\|\mathcal{C}_1 - \mathcal{C}_2\| = \|\mathcal{C}_1 \cup \{-\mathcal{C}_2\}\| = \sum_{k=1}^{K_1+K_2} \sum_{k'=1}^{K_1+K_2} \eta_k \cdot \eta_{k'} \kappa(c_k, c_{k'})$$

where K_1 and K_2 denote the number of faces in each of \mathcal{C}_1 and $-\mathcal{C}_2$ respectively.

We therefore define the currents-based metric for a pair of images $\mathcal{M}^{sc}(\mathcal{I}_i, \mathcal{I}_j)$ with extracted cortical surfaces $\mathcal{C}_i, \mathcal{C}_j$ as

$$\mathcal{M}^{sc}(\mathcal{I}_i, \mathcal{I}_j; \phi_{ij}) = \|\phi_{ij}(\mathcal{C}_i) - \mathcal{C}_j\|_c \quad (4)$$

One of the important benefits of the current metric is that there is no need for correspondence between the points of each mesh. This means that there is no requirement for the meshes to be re-parametrized based on, say, a template. The target mesh in Equation 4 is co-aligned to the source mesh using the transformation ϕ_{ij} before calculating the measure. The kernel width λ reflects the scale at which structures are compared and was set to $3mm$ for all comparisons.

C. Similarity- and distance-based measures: The effect of transformation scale

In broad terms, measures that relate pairs of images may be divided into similarity functions or distance functions. In the examples above, \mathcal{M}^a can be viewed as a similarity function and \mathcal{M}^{sd} and \mathcal{M}^{sc} can be viewed as distance functions. The scale of the transformations used to align the images, i.e. the degree to which local control is possible, the number of degrees of freedom in each transformation, has an effect on the pairwise measures and their power to provide suitable representations for the image data.

In the neonatal brain, changes in shape and appearance occur concurrently. If all the transformations are non-linear, aligning all pairs of images at a fine scale might be hoped to provide a means to separate these two effects. However, the registration process warps the images to optimise an intensity based similarity so that, in practice, fine scale alignment can also re-distribute signal patterns (as well aligning tissue boundaries), and thereby diminish signal differences. The global appearance measure \mathcal{M}^a for all comparisons will then tend to increase and may fail to characterise the fewer localised appearance differences that remain. As \mathcal{M}^a is bounded above, its dynamic range decreases and it therefore loses discriminative power.

On the other hand, if all the transformations are rigid, accounting only for differences in position and orientation between image pairs, the values of a shape measure such as \mathcal{M}^{sd} across all comparisons will become zero and, once again, discriminative power is lost. These considerations illustrate the fact that each pairwise measure for the image data has its own optimal transformation scale which leads to the best derived representation and which needs to be determined individually for each measure. For further discussion on the issues of transformation scale in the context of alignment and spatial normalization, readers are referred to [42], [43].

D. Deriving embeddings

Separate representations in the form of embedding coordinates are obtained by applying manifold learning algorithms to the pairwise measures of shape (\mathcal{M}^{sd} , \mathcal{M}^{sc}) and appearance (\mathcal{M}^a) among the images. We have chosen two established exemplars of manifold learning models designed for similarity- or distance-based data. We use the Laplacian eigenmap approach [15] as an example of the former and the Isomap algorithm [18] as an example of the latter and subsequently we evaluate all combinations of the different models with each type of data to test whether there is an effect on the final embeddings and to carry out a method selection step.

Using either approach, a weighted graph is used in which the nodes of the graph represent images and the weight associated with each edge indicates the proximity of the images at the nodes it joins. Edge weights are derived from the pairwise measures between the images.

Given a collection of images $\{\mathcal{I}_i : i = 1, \dots, N\}$, the Laplacian eigenmap approach uses an edge weight matrix $W = W_{ij}$ for the weights between images \mathcal{I}_i and \mathcal{I}_j and generates a set of embedding coordinates $\{y_i : i = 1, \dots, N\}$ in order to minimise the objective function

$$\mathcal{E}_{LE} = \sum_{i,j} W_{ij} \|y_i - y_j\|^2. \quad (5)$$

subject to a constraint to prevent the trivial solution where all y_i are set to zero. In Equation 5 it can be seen that, for higher values of W_{ij} , the objective function should generate correspondingly close coordinate pairs y_i and y_j hence the W_{ij} can be viewed as measures of affinity or similarity between the image pairs. In our work, we can either directly assign the appearance measure \mathcal{M}^a to W_{ij} or its value may be a similarity calculated from a shape measure such as \mathcal{M}^{sd} using a conversion function (see Section II-F). The norm $\|\cdot\|$ in Equation 5 is the L_2 norm and local structure in the data can be modelled by selecting a neighbourhood size K such that W_{ij} is set to zero for \mathcal{I}_j outside the K -nearest neighbourhood of \mathcal{I}_i , N_i^K , i.e.

$$\begin{aligned} W_{ij} &= \mathcal{M}(\mathcal{I}_i, \mathcal{I}_j; \phi_{ij}) && \text{for } \mathcal{I}_j \in N_i^K \\ &= 0 && \text{otherwise} \end{aligned} \quad (6)$$

where $\mathcal{M} = \mathcal{M}^a$ or $\mathcal{M} = f(\mathcal{M}^\theta)$ for some choice of conversion function f and $\theta \in \{sd, sc\}$.

The Isomap algorithm also makes use of edge weights but they are interpreted as distances \mathcal{D}_{ij} along the edges present in the graph. Once again, local structure is modelled by constructing a sparse graph in which edges are restricted to the K -nearest neighbourhood of each node. In other words, for the edges that are retained in the sparse graph, the values of \mathcal{D}_{ij} represent distances between pairs of ‘nearby’ images. In order to estimate the manifold structure, a full pairwise geodesic distance matrix $D = D_{ab}$ is estimated by identifying the shortest path between each distinct pair of images in the graph \mathcal{I}_a and \mathcal{I}_b for $a, b \in \{1, \dots, N\}$. D_{ab} is defined as the sum of $\{\mathcal{D}_{ai_1}, \mathcal{D}_{i_1i_2}, \dots, \mathcal{D}_{i_kb}\}$ which denote the edge distances along the shortest path between the images \mathcal{I}_a and \mathcal{I}_b in the graph.

Isomap then uses classical scaling [44] to find embedding coordinates $y_i, i = 1, \dots, N$ for the

data which minimise the cost function

$$\mathcal{E}_{Iso} = \sum_{i,j} (D_{ij} - \|y_i - y_j\|)^2 \quad (7)$$

so that the Euclidean distances between the learned coordinates match the estimated pairwise geodesic distances as closely as possible. In our work, we can either directly assign a shape measure, \mathcal{M}^{sd} or \mathcal{M}^{sc} , to \mathcal{D}_{ij} or its value may be calculated from the appearance measure \mathcal{M}^a .

A number of different methods may be used in order to optimise the objective functions defined in 5 and 7. Isomap and Laplacian eigenmaps both derive matrices from their graph representations. The eigenvalues and eigenvectors of these matrices are used to produce the embeddings which optimise the corresponding objective functions and this makes the algorithms examples of spectral methods of manifold learning or dimensionality reduction in relation to the mathematical sense of the term ‘spectrum’ to refer to the set of eigenvalues of a matrix. A good overview of spectral approaches may be found in [45].

E. Combining Embeddings

We distinguish between the different embedding coordinates y_i^{sd} , y_i^{sc} and y_i^a according to the type of measure used to generate them. The embedding y_i^θ for the i^{th} image where $\theta \in \{a, sd, sc\}$ is multivariate with dimension d^θ i.e. $y_i^\theta = (y_{i,1}^\theta, \dots, y_{i,d^\theta}^\theta)$ and the appropriate dimension d^θ needs to be determined in each case.

In the final step, we carry out a fusion of the embedding coordinates derived from shape and appearance measures, $\{y_i^{sd}\}$, $\{y_i^{sc}\}$ and $\{y_i^a\}$, $1 \leq i \leq N$. This is achieved by multiplying the zero-centred coordinates in each set by a scale factor so that we obtain unit variance for the first component in each set $y_{i,1}^\theta$, $\theta \in \{a, sd, sc\}$. After scaling, the separate embedding coordinates are concatenated to provide a $(d^a + d^{sd} + d^{sc})$ -dimensional coordinate set $\{Y_i = (f_a y_i^a, f_s y_i^{sd}, f_s y_i^{sc})\}$, $1 \leq i \leq N$. There is likely to be redundancy in the concatenated coordinates due to overlap in the descriptions of the image data given by the separate embeddings. This redundancy is then removed by using a further dimensionality reduction step applied directly to the coordinate set $\{Y_i\}$. The concatenated coordinates were used to provide a new set of pairwise distances using the L_2 distance, i.e. we calculate $\mathcal{D}_{ij} = \|Y_i - Y_j\|_2$ and a final Isomap step was then applied to the $\{\mathcal{D}_{ij}\}$ to give a combined coordinate embedding y_i^c with dimension d^c . We note, however, that the choice of dimension reduction method in the final combination step is less

crucial than the methods applied to the individual initial measure data. This is due to the fact that the input pairwise distances are already Euclidean as they are calculated using the L_2 norm between existing coordinate representations. This means that applying a method such as Isomap or MDS to the distances $\{Y_i\}$ becomes equivalent to the linear dimension reduction method of PCA and we have confirmed this experimentally for our data.

The approach above can be described as a ‘late-combination’ strategy for the different measure data because each measure is processed to produce its own set of embedding coordinates before combining them. An alternative approach would be to adopt an ‘early-combination’ strategy by generating a single pairwise measure from among the different types of measure data and using this in turn to produce a single set of embedding coordinates in one step.

F. Considerations on generating the weights matrices

Given the types of data and algorithms, conversion steps may be necessary for some combinations. When applying a Laplacian eigenmap step to the pairwise distance-based measures $\mathcal{M}^{sd/sc}$, a heat kernel function was used. Writing d_{ij} for the value of $\mathcal{M}^{sd/sc}(\mathcal{I}_i, \mathcal{I}_j; \phi_{ij})$, we make the assignment

$$W_{ij} = \exp\left(\frac{-d_{ij}^2}{\sigma^2}\right) \quad (8)$$

in order to construct the affinity matrix used in Equation 5. This use of a heat kernel allows the distances (low values for similar items) to be re-expressed as ‘affinities’ (high values for similar items) that are therefore suitable for use in algorithms such as Laplacian eigenmaps. The need for this type of conversion is due to the different assumptions made by differing algorithms over whether the input data are expressed as affinities or distances. General descriptions of heat kernels may be found in [46] and [45] and an example in the context of Laplacian eigenmaps may be found in [15]. When applying an Isomap step to pairwise similarity values among the images \mathcal{M}^a , the inverse of the function in Equation 8 is used. If s_{ij} denotes $\mathcal{M}^a(\mathcal{I}_i, \mathcal{I}_j; \phi_{ij})$, then we make the assignment

$$\mathcal{D}_{ij} = \sigma \sqrt{-2 \ln s_{ij}}. \quad (9)$$

Equations 8 and 9 effectively represent projections of the measures and the value of σ in each determines the location of the point of inflexion of the corresponding curves. The value of sigma was set to the median of the measure data in each case so that it corresponded with the densest

part of the data. As this point of inflection is located at the portion of the projection curve with the largest gradient magnitude, the densest part of the data corresponds to a wider range after projection in order to provide a greater level of differentiation among as many as possible of the different data item pairs.

When generating the weights matrices from the pairwise measures or derived functions of them, they are made symmetrical by averaging the initial calculated matrix with its transpose, i.e. we carry out the assignments

$$W_{ij} \leftarrow \frac{1}{2}(W_{ij} + W_{ji}) \quad \text{and} \quad \mathcal{D}_{ij} \leftarrow \frac{1}{2}(\mathcal{D}_{ij} + \mathcal{D}_{ji})$$

prior to solving the respective eigenproblems.

III. DATA AND PRE-PROCESSING

A three Tesla Philips Intera scanner was used to acquire 140 T2-weighted MR scans of 114 neonatal subjects with a range of ages. The parameters for the fast-spin echo (FSE) scans were: $TR = 17100\text{ ms}$, $TE = 160\text{ ms}$, $FA = 90^\circ$ and the resolution was $0.86 \times 0.86 \times 1\text{ mm}$. 69 scans were of female subjects and the gestational age information for the group, at birth and at scan, is given in table I. All images were visually inspected to ensure that they were free from motion artefact and major lesions. All images were corrected for MR inhomogeneity and the non-brain tissue was removed using ‘N3’ [47] and ‘BET’ [48].

TABLE I
AGE DATA FOR THE NEONATAL SUBJECTS STUDIED IN WEEKS OF GESTATIONAL AGE (GA).

	Mean	Standard Deviation	Minimum	Maximum
Birth	29.4	2.9	23.4	34.9
Scan	36.8	4.8	28.6	47.7

An illustration of scans obtained from subjects at various ages is given in Figure 1. As a pre-processing step to remove any global differences in size, all the scans were registered to a single chosen reference using 12 parameter affine transformations to account for position, orientation, scale and shear differences. The reference scan was chosen by a clinical expert (LS) as an image displaying an intermediate level of maturation within the cohort. The gestational age (GA) of the

reference scan subject was 32.0 weeks at birth and 36.3 weeks at scan. After affine registrations to the reference, all scans were aligned and resliced onto the same voxel grid.

The transformations $\{\phi_{ij}\}$ between all pairs of images were obtained using a registration model based on free-form deformations [49] (FFDs). The scale of each transformation is represented by the spacing of its set of control points with smaller spacings corresponding to a finer scale of local control. The registration model uses a multi-resolution scheme [50], which begins with a coarse control point grid and successively refines it. After carrying out all the registrations, transformations with control point spacings of 20, 10 and 5mm were available for each pair of images. For each image pair, and at control point spacing, registrations were carried out in both directions giving $140 \times 139 = 19460$ transformations at each scale.

Registrations between images and similarity comparisons were carried out using the ‘IRTK’ software package ¹. The Isomap [18] and Laplacian eigenmap [15] algorithms were implemented in ‘MatLab’ (Natick, MA). Statistical comparisons were carried out using MatLab except for comparisons of correlation coefficients which were carried out using ‘R’².

IV. RESULTS

We aim to produce coordinate embeddings for the image data with good representational power as measured by consistency with clinical data and self consistency. The next two sections describe how these were used for method selection among the combinations of manifold learning models and data and for choosing model parameters. In the subsequent sections, we present more detailed quantitative and visual comparisons of the embeddings produced by the selected method and the clinical data.

A. Method selection

Applying Isomap (*Iso*) and Laplacian eigenmaps (*LE*) to the chosen measures, \mathcal{M}^a , \mathcal{M}^{sc} and \mathcal{M}^{sd} gives eight possible method choices, each leading ultimately to its own set of combined embedding coordinates y^c . In order to select from among the method choices (e.g. $\mathcal{M}^a + Iso$ fused with $\mathcal{M}^{sd} + LE$ etc.), we assess the representational power of each y^c by its correlation

¹www.doc.ic.ac.uk/~dr/software

²www.r-project.org

with the ages of the subjects at time of scan and their ages at birth. The y_i^c were obtained up to 10 dimensions and the correlations of each component with the age data were calculated and the component with maximal correlation, ρ_{max} , was recorded.

In every case, ρ_{max} for the age at scan was always achieved by the first component, $y_{i,1}^c$, of the corresponding embedding and varied in the range $0.83 - 0.95$ (See Table II). The overall size of the brain is also highly correlated with age at scan ($\rho = 0.95$) but, as described in Section III, global size differences were removed prior to the analysis by affine registration of each scan to a reference. This is illustrated in Figure 5 which shows the brain volumes of the cohort before and after affine alignment to the reference and how, after alignment, most of the variation in brain volume due to age is removed. When comparing with the age at birth, ρ_{max} was achieved either for the second or the third component and varied in the range $0.37 - 0.68$ (See Table II).

To explore the possibility that the strongest associations with clinical data are not aligned with the coordinate axes of the learned embeddings, we also applied a canonical correlation analysis (CCA) in which one set of multivariate observations is represented by the combined age at scan and age at birth data and the second set is the learned embedding coordinates y_i^c for each method choice. CCA seeks a rotated basis in each observation space such that the first components in each basis are maximally correlated followed by the second and so on. The

TABLE II

MAXIMAL CORRELATIONS BETWEEN INDIVIDUAL COMPONENTS OF THE COMBINED EMBEDDING COORDINATES y_i^c AND EACH OF THE SCAN AND BIRTH AGE DATA. EACH ROW REPRESENTS A DIFFERENT ANALYSIS COMBINATION FOR THE SHAPE AND APPEARANCE MEASURES (SEE TEXT).

	Measure			ρ_{max}	
	\mathcal{M}^a	\mathcal{M}^{sd}	\mathcal{M}^{sc}	Scan age	Birth age
Method	Iso	Iso	Iso	0.95	0.61
	LE	Iso	Iso	0.95	0.68
	Iso	LE	Iso	0.95	0.47
	LE	LE	Iso	0.95	0.58
	Iso	Iso	LE	0.91	0.42
	LE	Iso	LE	0.89	0.40
	Iso	LE	LE	0.87	0.37
	LE	LE	LE	0.83	0.41

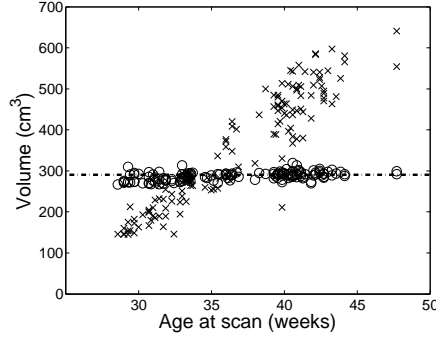


Fig. 5. The removal of global size differences in brain volume by affine alignment to a reference. This is carried out before deriving pairwise measures and learning embedding coordinates. The scatter graph shows brain volumes before alignment (crosses) and after alignment (circles) for the group plotted against age at scan. The dashed line shows the volume for the reference subject.

number of correlations is limited to two (as this is the dimension of the age data) and we denote them (ρ_1, ρ_2) . The values of each pair after CCA for each choice are shown in Table III.

TABLE III
CANONICAL CORRELATIONS BETWEEN THE EMBEDDING COORDINATES y_i^c AND THE COMBINED SCAN AND BIRTH AGE DATA TREATED AS A MULTIVARIATE OBSERVATION (SEE TEXT).

	Measure			Canonical correlations
	\mathcal{M}^a	\mathcal{M}^{sd}	\mathcal{M}^{sc}	(ρ_1, ρ_2)
Method	Iso	Iso	Iso	(0.96, 0.68)
	LE	Iso	Iso	(0.96, 0.70)
	Iso	LE	Iso	(0.96, 0.61)
	LE	LE	Iso	(0.96, 0.65)
	Iso	Iso	LE	(0.96, 0.65)
	LE	Iso	LE	(0.95, 0.65)
	Iso	LE	LE	(0.95, 0.62)
	LE	LE	LE	(0.95, 0.64)

All correlations in Tables II and III were significant at $p < 0.01$. The results in Table II indicate that all method choices provide strong associations with age at scan. The results in Table II right indicate a weaker association with age at birth particularly for choices of *LE* analysis on the shape data, \mathcal{M}^{sd} or \mathcal{M}^{sc} . The differences are less pronounced after rotating the

data under CCA as shown by the similarity of the columns in Table III. On the basis of these results, we selected the combination of $LE + \mathcal{M}^a$, $Iso + \mathcal{M}^{sc}$ and $Iso + \mathcal{M}^{sd}$ for subsequent analysis including further clinical data and this combination is used in all results below.

B. Choice of model parameters

Model parameters for our exemplar application to neonatal images were empirically determined based on the consistency of the input weights matrices with the embedding coordinates they lead to. This applies to all combinations of method choices described above. The model parameters to be determined are the scale of transformations between images, the dimensions of the intermediate and final combined embeddings and the value of k when applying $k - NN$ sparsification to the weights matrices. For the final choice of method, no conversions between similarities and distances were needed (Equations 8 and 9) so determining σ was not required.

The agreement between a $N \times N$ weight matrix A and a derived set of embedding coordinates y_i , $i = 1 \dots N$, was measured by the rank orderings each assigns to the input images \mathcal{I}_i . The matrix A either represents W for LE (Eqn. 5) or D for Isomap (Eqn. 7). In either case, the weights for image \mathcal{I}_i in A determine a rank ordering for all the images based on their proximity to \mathcal{I}_i . The embedding coordinates also provide an ordering for the data and Spearman's rank correlation may then be calculated for each image. The mean rank correlation $\bar{\rho}_{rank}$ across all images may then be used to represent the agreement of the weight matrix and the embedding coordinates. Each model parameter was varied to seek the best value of $\bar{\rho}_{rank}$ while keeping remaining parameters fixed. The final parameters were fixed after two passes with no significant change in value.

For the embedding dimensions, d^α , $\alpha \in \{a, sc, sd, c\}$, the values of $\bar{\rho}_{rank}$ were highest at $d^a = 5$ and reduced thereafter when using Laplacian eigenmaps. Using Isomap, the values of $\bar{\rho}_{rank}$ increase monotonically and asymptotically as the dimension increases but this can lead to over-fitting so we selected the minimal dimension for which $\bar{\rho}_{rank} > 0.7$ which lead to choices of $d^{sd} = 4$, $d^{sc} = 3$ and $d^c = 5$.

Using the same approach, the best scale of the transformations was an intermediate control point spacing of 10mm when obtaining the appearance embedding coordinates y_i^a , a coarser spacing (20mm) for the currents shape embedding y_i^{sc} and a finer spacing (5mm) was used for deriving the deformation shape embedding y_i^s . With the same approach, and searching in steps

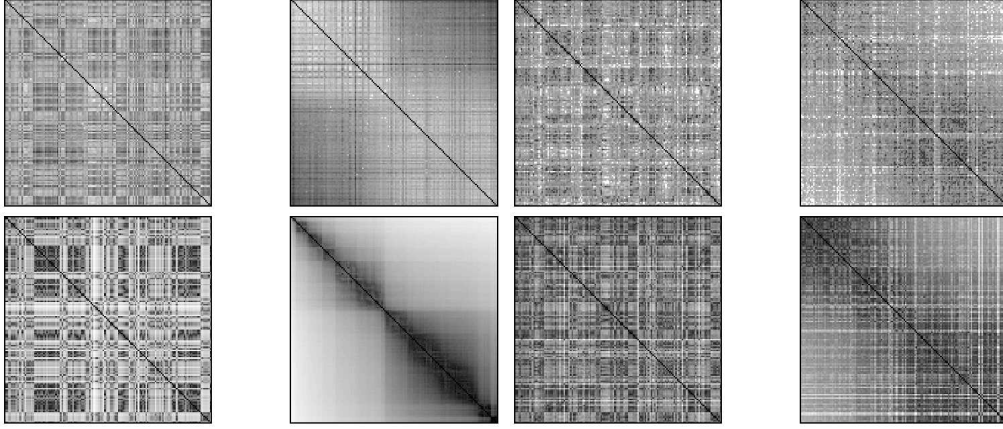


Fig. 6. Visualisations of the matrices used to generate the embedding coordinates. From top to bottom: \mathcal{M}^a , \mathcal{M}^{sd} , \mathcal{M}^{sc} and the L_2 distances between the concatenated coordinates. Left: the original matrices. Right: the matrices after re-ordering the rows and columns by the first component of the resulting embedding (see text).

of 5, the values of k used for sparsifying the weights matrices were 45 for y^a , 75 for y^{sd} , 40 for y^{sc} and 30 for the final combined embedding.

As an illustration of the process of converting input measure data into embedding coordinates, Figure 6 gives visualisations of the different measure matrices used. The matrices are shown in their original form (where the rows/columns are ordered by scan number Fig. 6 top) and after re-ordering the rows/columns by the value of the first component in the resulting embedding (Fig. 6 bottom). Taking the matrix for \mathcal{M}^a as an example (Fig. 6 top left), after re-ordering the higher similarity values (bright regions in Fig. 6 bottom left) tend to become closer to the main diagonal indicating greater affinity among ‘nearby’ scans in the embedding.

A comparison was made between the proposed late combination of separately derived embedding coordinates and early combination of the matrices prior to finding single embedding step. During early combination, both Laplacian eigenmaps and Isomap were tested and, in each case, the necessary conversions were made between distances and similarities (as described in Section II-F) prior to combining the matrices. The matrices were combined by first scaling them to unit mean and then adding them in linear combination. This necessitated two parameters for the weights, i.e. a single weights matrix is obtained from three separate matrices, $M_{1,2,3}$ as $M = \beta M_1 + \gamma M_2 + (1 - \beta - \gamma) M_3$. For each of the Isomap and Laplacian Eigenmap applications,

TABLE IV

OBSERVED CORRELATIONS BETWEEN CLINICAL DATA AND THE FIRST COMPONENTS OF THE INTERMEDIATE (y^{sd} , y^{sc} , y^a) AND COMBINED EMBEDDING COORDINATES (y^c). GESTATIONAL AGE (GA), WEIGHT (W) AND HEAD CIRCUMFERENCE (HC) DATA WERE OBTAINED AT SCAN AND AT BIRTH.

		$y_{i,1}^c$	$y_{i,1}^{sd}$	$y_{i,1}^{sc}$	$y_{i,1}^a$	$y_{i,2}^c$	$y_{i,2}^{sd}$	$y_{i,2}^{sc}$	$y_{i,2}^a$
Scan	GA	0.95*	0.88*	0.92*	0.92*	0.02	0.07	0.10	0.03
	W	0.92*	0.85*	0.90*	0.89*	0.04	0.08	0.12	0.04
	HC	0.95*	0.88*	0.89*	0.88*	0.05	0.18	0.19	0.07
Birth	GA	0.20	0.21	0.22	0.07	0.68*	0.61*	0.26†	0.49*
	W	0.25	0.24	0.26	0.17	0.48*	0.45*	0.14	0.35†
	HC	0.27	0.26	0.30	0.16	0.54*	0.56*	0.20	0.35†

[* $p < 10^{-4}$, † $p < 10^{-3}$]

the combinations of β and γ in $\{0, 0.1, \dots, 1\}$ were tested to find the maximal correlation with the non-imaging data.

C. Comparison with clinical information

Various data for the subjects were collected at birth and at time of scan. Apart from the ages at birth and scan, measurements of head circumference and weight were also available as well as the sex of each subject. For each of the intermediate shape and appearance embedding coordinates, y_i^{sd} , y_i^{sc} and y_i^a and for the combined embedding y_i^c of the N images, the explanatory power was assessed based on the corresponding correlations with the clinical data. Associations with age at scan are of interest as it is a major factor in the morphology of the brain in the period of rapid neonatal growth. The age at birth is also of interest as it is indicative of the degree of prematurity.

The correlations obtained are listed in Table IV in which it can be seen that the first component, $y_{i,1}^\alpha$, $\alpha \in \{a, sd, sc, c\}$, in every case has a strong correlation with clinical data obtained at scan. Additionally, weaker correlations which mostly remained significant were observed between the second components in each case - $y_{i,2}^\alpha$, $\alpha \in \{a, sd, sc, c\}$ - and the clinical data obtained at birth. The components from the currents shape metric gave the weakest correlations with age at birth among the separate embeddings.

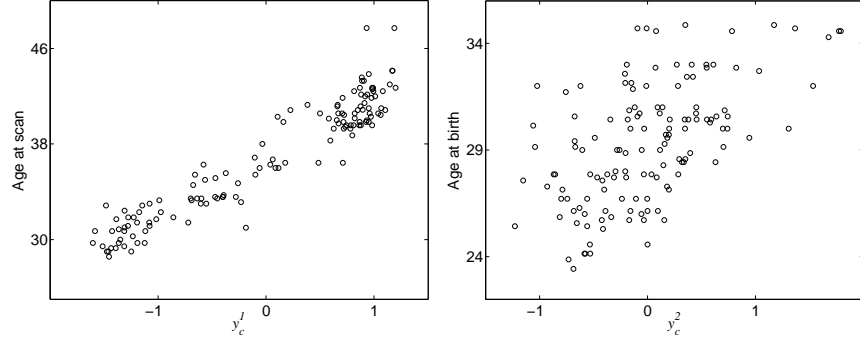


Fig. 7. The first two combined embedding components, each plotted against the different age data. Top: $y_{i,1}^c$ against weeks of gestational age (GA) at scan. Bottom: $y_{i,2}^c$ against GA at birth.

The size of the subject at time of scan, in our case reflected by head circumference and weight, will clearly play an important part in characterising the acquired scan. However, as noted earlier, differences in global size for all subjects were removed prior to subsequent processing using 12 parameter affine transformations (See Figure 5). The correlations observed in the top three rows of Table IV were obtained from images that had been affinely co-aligned.

Lines have been drawn in Table IV to emphasise the block structure of the correlation values. The entries can be seen as a 2×2 block matrix where the diagonal elements are high and the off-diagonal elements are low. This suggests orthogonality among the correlations between the first and second embedding components on one hand and the data acquired at birth and scan on the other.

Focusing on the diagonal blocks, the benefit of combining the separate embeddings into a single combined embedding may be seen. In nearly all cases, correlations for the combined embedding exceed the corresponding correlations for the separate embeddings with one exception in the final row. The significance of the increase in age data correlation for the combined embedding against the single embeddings (rows 1 and 4 in Table IV) were estimated using the method of Cohen & Cohen [51] for comparing dependent correlations (as implemented in R). The differences were all significant at $p < 0.1$ except for the comparison between $\rho(y_{i,2}^c, \text{Birth GA})$ (0.68) and $\rho(y_{i,2}^{sd}, \text{Birth GA})$ (0.61, $p = 0.06$).

On considering further components beyond those shown in Table IV, weak but significant correlations between each of $y_{i,5}^s$ and $y_{i,5}^c$ and the sex of the subjects were observed: 0.24 ($p <$

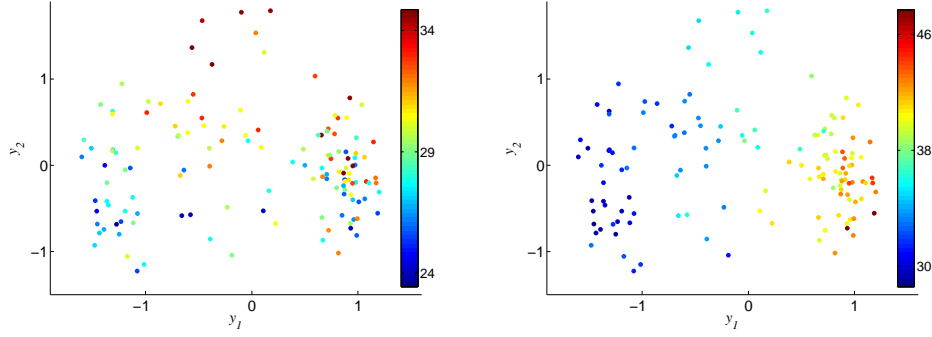


Fig. 8. A scatter graph of combined embedding components $y_{i,1}^c$ and $y_{i,2}^c$. The colour coding indicates age at birth (top) and age at scan (bottom).

0.02) and 0.27 ($p < 0.01$) respectively. Dividing the subjects according to sex and applying a 2-sided unequal variance t-test based on the values of $y_{i,5}^c$ gave a t-statistic of 2.85 ($p < 0.01$).

As a comparison, the single embedding obtained after the optimal early-combination of the data into a single measure matrix was 0.94 against age at scan for the first component and 0.61 against age at birth for the second component. The optimal method for the early combination strategy was Isomap and the weights used were 0.5, 0.4 and 0.1 for the matrices \mathcal{M}^a , \mathcal{M}^{sd} and \mathcal{M}^{sc} . The differences between these correlations and the corresponding correlations for $y_{1,2}^c$ obtained by late combination (in Table IV) was significant for age at birth but not significant for the age at scan ($p < 0.01$).

D. Visualising the embeddings

The relationships between each of the first two components of y_i^c and the ages at scan and at birth are shown in scatter plots in Figure 7. A visualisation of the manifold structure of the data in the form of a plot of the first two components of y_i^c against each other is given in Figure 8. This plot is given twice with different colour codings, one to indicate age at birth for the subject and one to indicate age at scan. Similar scatter plots are shown for each of the separate input embedding coordinates in Figure 9, colour-coded to indicate age at scan.

The scatter plot of the first two embedding coordinates is also shown using slices taken from the image data at the level of the ventricles in Figure 10 to illustrate the changes in overall shape and appearance of the data in the space of the learned coordinates.

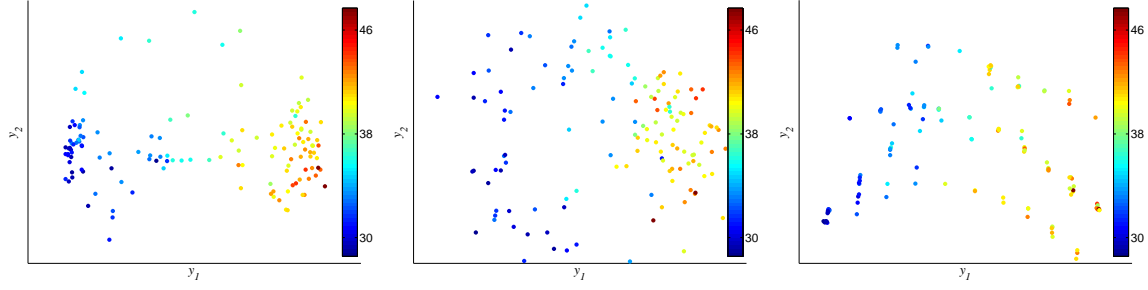


Fig. 9. Scatter graphs to illustrate the first two components of each of the separate input embeddings prior to combination (see figure 8 for the combined embedding coordinates). Colour coding indicates the age at scan.

Figures 8 and 9 present a visualisation of the main components of each of input and combined embeddings in relation to the age data, indicating the strength of the relationship amongst them. The input embeddings are different from each other, as shown in Figure 9, and each differs from the combined embedding. The pattern of morphology in Figure 10 generally suggests a change from simpler to more complex morphology from left to right. The change in morphology in the vertical direction is not clear despite the correlation with age at birth and suggests a possible direction for further research.

The ability to represent the image data using a set of embedding coordinates enables further types of analysis to be made. A good example is the delineation of growth trajectories through the image data using kernel smoothing, as in the method used by Hutton et al. [52] for shape space data. The criteria defining trajectories may be varied and they allow the data to be ‘navigated’ in different ways. If, for example, a covariate t is represented by age at scan, a continuous trajectory $y^c(t)$ through the combined embedding coordinates may be parametrised by t using a kernel smoothing expression

$$y^c(t) = \frac{1}{\nu} \sum_{i=1}^N K(t_i - t) y_i^c$$

where t_i is the scan age of the subject in image \mathcal{I}_i , K is a smoothing kernel and ν is a normalising constant. Using a Gaussian kernel, an example of such a trajectory, after projection onto the first two components, is shown in Figure 11(a). By selecting a number of equally spaced points along the trajectory, reconstructions of images at their locations may be estimated from the image data set based on their distance in embedding coordinates. This again may be carried out using a

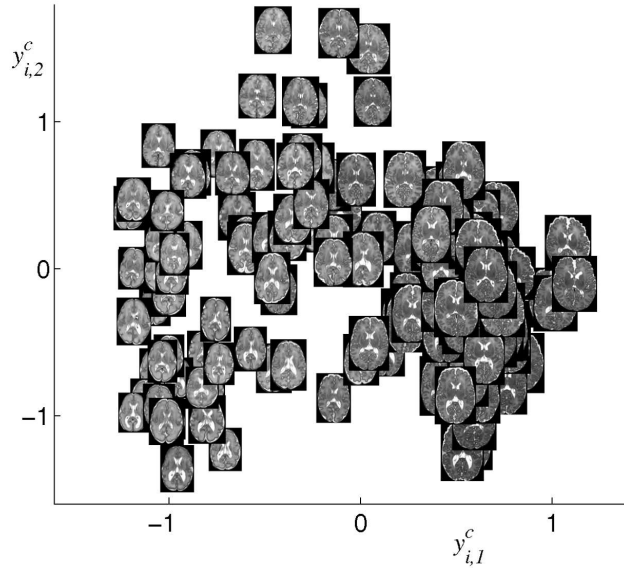


Fig. 10. An illustration of the variation in morphology of the data using slices from the image data arranged with the same embedding coordinates illustrated in Figure 8

kernel regression step for image averaging, further details of which may be found Ericsson et al. [53] or Gerber et al. [17]. Figure 11(b) shows reconstructed images at ten equally spaced points on the trajectory in 11(a) as an example. Note that the reconstructions are also normalised for global size as they were reconstructed from the pre-processed image data.

V. DISCUSSION

The framework we have proposed combines the results of separate manifold learning steps into a single and more powerful representation of an image dataset that is useful in the context of population studies. Each of the separate steps analyses its own pairwise measure relevant to specific morphological aspects of the clinical and biological context of the image data. The ability to generate good representations of such populations of subjects is important as it forms a step towards biomarker discovery which, in the case of our example application, should relate to patterns of neonatal brain development. While the example we have presented relates to neonatal brain image data, the framework for combining embeddings is a general one and in principle it may be applied to other cohorts and types of imaging data.

As the example we present focuses on the early development of the neonatal brain, the shape

and MR appearance are relevant and were used as a basis of pairwise measures derived either directly from the deformations between images or from the intensity based similarities they induce. We emphasise the importance of the choice of pairwise measure between images to the clinical context of the data and the biological features of interest. A similar general approach in the context of non-clinical data is given in the work of Souvenir and Pless [54] who characterise the various types of transformation that can relate images based on a suitably defined pairwise image distance measure for each.

In our example application, pairwise measures were derived directly or indirectly from non-rigid deformations between image pairs. We found that the optimal transformation scale for each type of measure varied, with a fine scale needed for the shape measure (deformation based) and an intermediate scale for the appearance measure (image intensity based). These issues regarding transformation scale are analogous to similar issues in other morphometric studies. In the case of voxel-based morphometry (VBM) [1], local differences between groups are identified in sets of co-aligned images [2], [3]. The transformation scale used when aligning the data should preserve the detectability of the group differences. If, however, a very fine-scale co-alignment is applied, a larger proportion of the differences among the images becomes encoded in the deformations between them. In this case a study using the approach of deformation-based morphometry (DBM) becomes more appropriate [4]–[6].

The representational power of the derived embeddings was assessed according to their correlation with clinical data relating to the subjects scanned, such as their ages. During the method selection step, varying the combinations of pairwise measure and different manifold learning technique led to some variation in the representational power indicating a degree of robustness across methods. We observed an advantage for the application of Isomap to the shape measures and a slight advantage in applying Laplacian embedding to the appearance measure (Tables II). This may be due to the absence of an extra step to convert measures between similarity and distance functions but further work would be required to establish this. The model parameter selection (Section IV-B) was data driven according to the agreement of input measures and embeddings by rank correlation. In the case of the nearest neighbour counts for each of the embedding steps, this led to values that were fairly high with respect to the size of the data set. Developing a more tailored scheme for finding the optimal neighbourhood size may form a useful focus for future work. Methods for finding appropriate neighbourhood sizes tend to

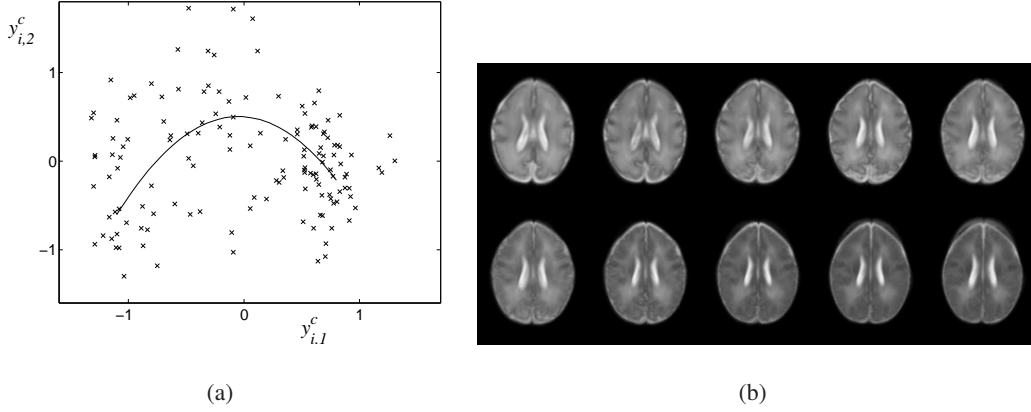


Fig. 11. (a) A kernel smoothing trajectory in the embedding coordinates of the image data. (b) Reconstructed images along the trajectory at ten equally spaced locations. These images retain the global size normalisation applied during pre-processing.

be heuristic (See for example [45], Sec. 8.1.3) and we are not aware of an established analytic method for carrying this out in the general machine learning literature.

The patterns of correlation with clinical data (Table IV) are broadly similar among the separate shape and appearance embeddings and the final combined embedding which suggests a degree of overlap in the descriptions they provide. The consistency across embeddings along with the strong correlations with clinical information suggests, to some extent, a degree of biological plausibility to the generated representation.

We combine the separate embedding coordinates by concatenation and further dimension reduction. While overlap exists in the descriptions provided by the separate embeddings, some independence of description is also evidenced by the improved correlations obtained after combining them. An alternative approach would be to combine the measure matrices prior to the application of the manifold learning step in a way similar to the combination of kernel data in kernel machine learning methods (See e.g. [46]). When applying such an early combination strategy to our example data, we obtained a lower correlation with scan age than that obtained under late combination (0.94 vs. 0.96) and a significantly lower correlation with birth age data (0.61 vs 0.68). This suggests some benefit to the late combination strategy proposed which may be due to the ability to select and apply different manifold learning techniques and parameters for each measure. This means that each input measure may undergo a different mapping to its optimal intermediate embedding coordinates prior to combination and, in our case, the mappings

used (Isomap, LE) are non-linear.

In all cases, the first component of each set of embedding coordinates correlated strongly with the age at scan. This result is expected as the age at scan is the dominant factor governing the shape of the brain and its appearance in an MR image. We also expect that correlations of embedding components with size data, such as weight and head circumference, should parallel correlations with the corresponding age (at birth or at scan). This is because the size measurements themselves are correlated with age but the results at least provide further proof-of-principle for the method in general. As stated previously, these results were obtained after a pre-processing step to correct for affine global variations in size. In our example application, we have also observed orthogonality among the correlations (Table IV) which indicates an independence among the descriptions provided by first two components in each embedding.

In the example application presented, the use of affine spatial normalisation as a pre-processing step made the subsequent analysis using non-linear transformations much simpler. This is because each of the inter-scan transformations used could be defined in the same space (the space of the single template) and measures defined from them would be commensurable. It may be interesting in future work, however, to explicitly incorporate affine transformations between scans into the process of finding measures. This would have the effect of incorporating global size and shape information into the analysis in common with previous volumetric studies of neonatal data ([29] for example).

We note with interest that most of the second components of the embeddings tend to correlate with age at birth as this suggests that, after age at scan, variation in brain morphology may be accounted for by age at birth. This coincides with earlier work on the possibly lasting effect of prematurity on brain development [55]. In comparison with our earlier work [36], the incorporation of the cortex-derived shape measure data resulted in a stronger association of the combined embedding with age at scan but an unchanged correlation with age at birth. It is predictable that the shape measure should be most strongly associated with the age at scan because of the rapid growth of the brain over this period, expressing the precise neurobiological process of brain development. However the embeddings also show an effect of the age at birth, suggesting that the environmental effect of premature extrauterine life disrupts the genetic programme of brain growth, and thus the embeddings reveal genetic and environmental effects on brain development. The appearance measure probably also detects these effects, with the

difference that the environmental effects on appearance are the most obvious; this is consistent with our previous demonstration that prematurity commonly leads to diffuse excessive high signal in white matter on T2 weighted images. These observations of the disruption of the genetic programme of brain development by the environmental effect of prematurity are generally consistent with the changes in structure and microstructural appearance that we have shown using Deformation Based Morphometry and Tract Based Spatial Statistics [56].

We have also shown an example of the type of embedding coordinate trajectory that may be defined in order to navigate the data. We illustrate how this may be carried out by applying kernel regression to the coordinate representation and also that kernel regression may also be used to reconstruct images at locations along the trajectory. We have given an example in which the regression covariate was the age at scan although many other clinical or non-clinical covariates are possible. The ability to characterise images along a trajectory clearly has significant potential for our application to the development of the neonatal brain by, for example, distinguishing between normal and abnormal patterns of growth. More generally, the ability to characterise changes along a trajectory defined by a given covariate is a potentially powerful tool in the context of biomarker discovery. The reconstruction of images along an age trajectory is an example of representations that may be derived from the quantitative features, i.e. the embedding coordinates, given by the manifold learning step. This application contrasts with the methods of, for example Habas et al. [28] and Murgasova et al. [29] which provide important qualitative snapshots of development directly from the image data, and our quantitative embedding coordinates, provided by combined measure manifold learning, may potentially be directly applied in a wider range of contexts, such as clustering or classification.

One of the main limitations of the proposed framework lies in the amount of processing required; the number of comparisons, for example, grows quadratically with the number of images and the required non-rigid registrations may be computationally costly to obtain. Heuristic approaches for reducing the work-load have however, been applied, for example Gerber et al. [17] curtail registration optimisations between image pairs if they run beyond a threshold number of iterations and the corresponding pair of images are deemed to be outside each other's nearest neighbourhoods. Another important limitation lies in the possible applications for the representations given by non-linear manifold learning techniques. For example, if the representation is used to train a classifier (See e.g., [57]) then it may be important to be able

to process new test data in a manner in which the bulk of the training is ‘offline’. As manifold learning algorithms typically require all the data to derive a representation, this may be difficult and contrasts with subspace learning methods such as PCA in which a single linear projection can be learned once and subsequently applied to all new data.

Alternative projection methods have been proposed, for example the kernel regression method for reconstructing images from embedding coordinates (Section IV-D) may be viewed as a projection from the embedding coordinate domain to the image domain. A similar kernel regression approach for carrying out the projection from images to embedding coordinates as a means for dealing with new data is given by Gerber et al. [17]. He et al. [25] proposed learning an optimal linear projection that preserves the pairwise measures in the context of face recognition although this may incur a loss of the flexibility provided by non-linear representations. If all the data is to be re-processed when a new image is given, then this may affect the representation provided and its performance on the application used. In an application of spectral clustering for clinical classification we were previously able to show that the embedding coordinates are reasonably stable if the size of the training data set is sufficiently large [16].

Finally, there are a number of other interesting directions in which the work presented may be developed. One example is to analyse local structure more directly by making pairwise measures relate to specific regions rather than defining them as global measures between scans as we have done. The effect of scale can also be investigated by deriving measures using a multi-resolution framework, either of the images in the form of, say, image pyramids, or of the transformations by incorporating measures from multiple scales of transformation simultaneously. In terms of clinical data, it is clearly possible to explicitly incorporate this into the derivation of representations in a supervised way. In this work, we have generated new representations based on image-derived measures and compared them with clinical data post-hoc. The alternative would be to derive new coordinate representations from a combination of clinical data and image-derived measures. Such representations may be used to create, for example, classifiers or regressors for clinical data or biomarkers.

REFERENCES

- [1] J. Ashburner and K. J. Friston, “Voxel-based morphometry – the methods,” *NeuroImage*, vol. 11, no. 6, pp. 805–821, 2000.
- [2] C. Good, I. Johnsrude, J. Ashburner, R. Henson, K. Friston, and R. Frackowiak, “A voxel-based morphometric study of ageing in 465 normal adult human brains,” *NeuroImage*, vol. 14, no. 1, pp. 21–36, 2001.

- [3] M. Kubicki, M. E. Shenton, D. F. Salisbury, Y. Hirayasu, K. Kasai, R. Kikinis, F. A. Jolesz, and R. W. McCarley, "Voxel-based morphometric analysis of gray matter in first episode schizophrenia," *NeuroImage*, vol. 17, no. 4, pp. 1711–1719, 2002.
- [4] C. Gaser, I. Nenadoc, B. R. Buchsbaum, E. A. Hazlett, and M. S. Buchsbaum, "Deformation-based morphometry and its relation to conventional volumetry of brain lateral ventricles in MRI," *NeuroImage*, vol. 13, pp. 1140–1145, 2001.
- [5] T. Rohlfing, E. V. Sullivan, and A. Pfefferbaum, "Deformation-based brain morphometry to track the course of alcoholism: Differences between intra-subject and inter-subject analysis," *Psychiatry Research: Neuroimaging*, vol. 146, no. 2, pp. 157–170, 2006.
- [6] C. Studholme, "Incorporating DTI data as a constraint in deformation tensor morphometry between T1 MR images," in *Information Processing in Medical Imaging: Proc. 20th International Conference (IPMI'07)*, ser. Lecture Notes in Computer Science, vol. 4584. Springer, 2007, pp. 223–232.
- [7] Z. Lao, D. Shen, Z. Xue, B. Karacali, S. M. Resnick, and C. Davatzikos, "Morphological classification of brains via high-dimensional shape transformations and machine learning methods," *NeuroImage*, vol. 21, no. 1, pp. 46–57, 2004.
- [8] Y. Fan, N. Batmanghelich, C. M. Clark, and C. Davatzikos, "Spatial patterns of brain atrophy in MCI patients, identified via high-dimensional pattern classification, predict subsequent cognitive decline," *NeuroImage*, vol. 39, pp. 1731–1743, 2008.
- [9] S. Baloch and C. Davatzikos, "Morphological appearance manifolds in computational anatomy: Groupwise registration and morphological analysis," *NeuroImage*, vol. 45, no. 1, pp. S73–S85, 2009, mathematics in Brain Imaging.
- [10] F. Shi, P. T. Yap, Y. Fan, J. H. Gilmore, W. Lin, and D. Shen, "Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation," *NeuroImage*, vol. 51, no. 2, pp. 684–693, 2010.
- [11] Y. Wang, Y. Fan, P. Bhatt, and C. Davatzikos, "High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables," *NeuroImage*, vol. 50, no. 4, pp. 1519–1535, 2010.
- [12] L. van der Maaten, E. O. Postma, and H. van den Herik, "Dimensionality reduction: A comparative review," Tilburg University, NL, Tech. Rep. TiCC-TR 2009-005, 2009. [Online]. Available: http://homepage.tudelft.nl/19j49/Publications_files/TR_Dimensiereductie.pdf
- [13] G. Langs and N. Paragios, "Modeling the structure of multivariate manifolds: Shape maps," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR'08)*. IEEE, 2008.
- [14] R. R. Coifman and S. Lafon, "Diffusion maps," *Applied and Computational Harmonic Analysis*, vol. 21, no. 1, pp. 5–30, 2006.
- [15] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [16] P. Aljabar, D. Rueckert, and W. R. Crum, "Automated morphological analysis of magnetic resonance brain imaging using spectral analysis," *NeuroImage*, vol. 43, no. 2, pp. 225–235, 2008.
- [17] S. Gerber, T. Tasdizen, P. T. Fletcher, S. Joshi, and R. Whitaker, "Manifold modeling for brain population analysis," *Medical Image Analysis*, vol. 14, no. 5, pp. 643–653, 2010.
- [18] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [19] J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos, "GRAM: A framework for geodesic registration on anatomical manifolds," *Medical Image Analysis*, vol. 14, no. 5, pp. 633–642, 2010.

- [20] R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, D. Rueckert, and the Alzheimer's Disease Neuroimaging Initiative, "LEAP: Learning embeddings for atlas propagation," *NeuroImage*, vol. 49, no. 2, pp. 1316–1325, 2010.
- [21] Z. Y. Sun, M. Perrot, A. Tucholka, D. Rivière, and J. F. Mangin, "Constructing a dictionary of human brain folding patterns," in *Twelfth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '09)*, ser. Lecture Notes in Computer Science, vol. 5762. Springer, 2009, pp. 117–124.
- [22] S. Atasoy, D. Mateus, J. Lallemand, A. Meining, G. Z. Yang, and N. Navab, "Endoscopic video manifolds," in *Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)*, ser. Lecture Notes in Computer Science, vol. 6362. Springer, 2010, pp. 437–445.
- [23] C. Wachinger, M. Yigitsoy, and N. Navab, "Manifold learning for image-based breathing gating with application to 4D ultrasound," in *Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)*, ser. Lecture Notes in Computer Science, vol. 6362. Springer, 2010, pp. 26–33.
- [24] R. Sparks and A. Madabhushi, "Novel morphometric based classification via diffeomorphic based shape representation using manifold learning," in *Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)*, ser. Lecture Notes in Computer Science, vol. 6363. Springer, 2010, pp. 658–665.
- [25] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 328–340, 2005.
- [26] J. Zhang, S. Li, and J. Wang, "Manifold learning and applications in recognition," in *Intelligent Multimedia Processing with Soft Computing*, ser. Studies in Fuzziness and Soft Computing, vol. 168, 2005, pp. 281–300.
- [27] F. Yun and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *Multimedia, IEEE Transactions on*, vol. 10, no. 4, pp. 578–584, 2008.
- [28] P. A. Habas, K. Kim, J. M. Corbett-Detig, F. Rousseau, O. A. Glenn, A. James Barkovich, and C. Studholme, "A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation," *NeuroImage*, vol. 53, no. 2, pp. 460–470, 2010.
- [29] M. Murgasova, P. Aljabar, L. Srinivasan, S. J. Counsell, V. Doria, A. Serag, I. S. Gousias, J. P. Boardman, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "A dynamic 4D probabilistic atlas of the developing brain," *NeuroImage*, vol. 54, no. 4, pp. 2750–2763, 2011.
- [30] F. Shi, Y. Fan, S. Tang, J. H. Gilmore, W. Lin, and D. Shen, "Neonatal brain image segmentation in longitudinal MRI studies," *NeuroImage*, vol. 49, no. 1, pp. 391–400, 2010.
- [31] P. A. Habas, K. Kim, F. Rousseau, O. A. Glenn, A. J. Barkovich, and C. Studholme, "Atlas-based segmentation of developing tissues in the human brain with quantitative validation in young fetuses," *Human Brain Mapping*, vol. 31, no. 9, pp. 1348–1358, 2010.
- [32] M. Prastawa, N. Sadeghi, J. H. Gilmore, W. Lin, and G. Gerig, "A new framework for analyzing white matter maturation in early brain development," in *IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, 2010*, 2010, pp. 97–100.
- [33] A. Serag, P. Aljabar, J. V. Hajnal, S. J. Counsell, J. P. Boardman, and D. Rueckert., "Tracking developmental changes in subcortical structures of the preterm brain using multi-modal MRI." in *IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, 2011*, 2011.
- [34] M. Sabuncu, S. Balci, M. Shenton, and P. Golland, "Discovering modes of an image population through mixture modeling," in *Eleventh Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '08)*, ser. Lecture Notes in Computer Science, vol. 5242, 2008, pp. 381–389.

- [35] D. Blezek and J. Miller, "Atlas stratification," *Medical Image Analysis*, vol. 11, no. 5, pp. 443–457, 2007.
- [36] P. Aljabar, R. Wolz, L. Srinivasan, S. J. Counsell, J. P. Boardman, M. Murgasova, V. Doria, M. A. Rutherford, A. D. Edwards, J. V. Hajnal, and D. Rueckert, "Combining morphological information in a manifold learning framework: Application to neonatal MRI," in *Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)*, ser. Lecture Notes in Computer Science, vol. 6363, 2010, pp. 1–8.
- [37] M. Vaillant and J. Glaunés, "Surface matching via currents," in *Information Processing in Medical Imaging: Proc. 19th International Conference (IPMI'05)*, ser. Lecture Notes in Computer Science, vol. 3565, 2005, pp. 381–392.
- [38] M. F. Beg, M. I. Miller, A. Trounev, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139–157, 2005.
- [39] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [40] L. Ha, M. Prastawa, G. Gerig, J. H. Gilmore, C. T. Silva, and S. Joshi, "Image registration driven by combined probabilistic and geometric descriptors," in *Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI '10)*, ser. Lecture Notes in Computer Science, vol. 6362. Springer, 2010, pp. 602–609.
- [41] H. Xue, L. Srinivasan, S. Jiang, M. Rutherford, A. D. Edwards, D. Rueckert, and J. V. Hajnal, "Automatic segmentation and reconstruction of the cortex from neonatal MRI," *NeuroImage*, vol. 38, no. 3, pp. 461–477, 2007.
- [42] W. R. Crum, L. D. Griffin, D. L. G. Hill, and D. J. Hawkes, "Zen and the art of medical image registration: correspondence, homology, and quality," *NeuroImage*, vol. 20, no. 3, pp. 1425–1437, 2003.
- [43] E. Haber and J. Modersitzki, "COFIR: coarse and fine image registration," in *Real-Time PDE-Constrained Optimization*. SIAM, 2006, pp. 37–49.
- [44] W. Torgerson, "Multidimensional scaling: I. Theory and method," *Psychometrika*, vol. 17, pp. 401–419, 1952.
- [45] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [46] J. Shawe-Taylor and N. Cristianini, Eds., *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004.
- [47] J. G. Sled, A. P. Zijdenbos, and A. C. Evans, "A non-parametric method for automatic correction of intensity non-uniformity in MRI data," *IEEE Transactions on Medical Imaging*, vol. 17, no. 1, pp. 87–97, 1998.
- [48] S. M. Smith, "Fast robust automated brain extraction," *Human Brain Mapping*, vol. 17, no. 3, pp. 143–155, Nov 2002.
- [49] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes, "Non-rigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [50] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truweit, F. A. Gerritsen, D. L. G. Hill, and D. J. Hawkes, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Fourth Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI '01)*, ser. Lecture Notes in Computer Science, W. J. Niessen and M. A. Viergever, Eds., vol. 2208, 2001, pp. 573–581.
- [51] J. Cohen and P. Cohen, *Applied multiple regression/correlation analysis for the behavioral sciences*. Lawrence Erlbaum, Hillsdale NJ, 1983.
- [52] T. J. Hutton, B. F. Buxton, P. Hammond, and H. Potts, "Estimating average growth trajectories in shape-space using kernel smoothing," *IEEE Transactions on Medical Imaging*, vol. 22, no. 6, pp. 747–753, 2003.
- [53] A. Ericsson, P. Aljabar, and D. Rueckert, "Construction of a patient-specific atlas of the brain: Application to normal

- aging,” in *IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, Paris, April 2008*, 2008, pp. 480–483.
- [54] R. Souvenir and R. Pless, “Image distance functions for manifold learning,” *Image Vision Computing*, vol. 25, no. 3, pp. 365–373, 2007.
- [55] J. J. Volpe, “The encephalopathy of prematurity – Brain injury and impaired brain development inextricably intertwined,” *Seminars in Pediatric Neurology*, vol. 16, no. 4, pp. 167–178, 2009.
- [56] G. Ball, P. J. Boardman, D. Rueckert, P. Aljabar, T. Arichi, N. Merchant, I. S. Gousias, A. D. Edwards, and S. J. Counsell, “The effect of preterm birth on thalamic and cortical development,” *Cerebral Cortex*, vol. (In press), 2011.
- [57] R. Wolz, P. Aljabar, J. V. Hajnal, J. Lötjönen, and D. Rueckert, “Improved manifold classification incorporating subject knowledge,” in *IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro, 2011*, 2011, pp. –.